

CLAIMS

What is claimed is:

1 1. A method of operation on one or more data processing machines, the
2 method comprising:
3 determining a first collection rating for a first rating scale for contents of a first
4 document collection;
5 determining a first link rating for said first rating scale for contents linked to or
6 linked by contents of said first document collection; and
7 modifying said first collection rating for said first rating scale for contents of
8 said first document collection based on said determined first link rating for said first
9 rating scale for contents linked to or linked by contents of said first document
10 collection.

1 2. The method of claim 1, wherein said determining of a first collection rating
2 comprises determining said first collection rating based on document ratings of a
3 first subset of documents of said first collection of documents, and sizes of the
4 documents of the first subset of documents of the first document collection.

1 3. The method of claim 2, wherein said first subset of documents of said first
2 document collection consists of first textual documents of said first document
3 collection.

1 4. The method of claim 1, wherein said determining of a first link rating
2 comprises determining at least a second collection rating for at least a second

3 document collection with documents linked to or linked by documents of said first
4 document collection, and determining said first link rating based on said determined
5 at least a second collection rating of said at least a second document collection.

1 5. The method of claim 1, wherein said modifying of the first collection rating
2 comprises replacing the determined first collection rating with said determined first
3 link rating.

1 6. The method of claim 1, wherein said modifying of the first collection rating
2 comprises adding said determined first link rating to the determined first collection
3 rating.

1 7. The method of claim 1, wherein said modifying of the first collection rating
2 comprises subtracting said determined first link rating from the determined first
3 collection rating.

1 8. The method of claim 1, wherein said first document collection is a web site,
2 and said contents of said first document collection are web pages.

1 9. A method of operation on one or more data processing machines, the
2 method comprising:
3 determining document ratings for a rating scale for a subset of documents of
4 a document collection;
5 determining sizes of the documents of said subset;

6 determining a collection rating for said rating scale for said document
7 collection based on said determined document ratings of said subset of documents,
8 and normalized by said determined sizes of said subset of documents.

1 10. The method of claim 9, wherein said determining of the collection rating
2 comprises further subdividing said subset of documents into a plurality of groups in
3 accordance with their determined sizes, and applying a weight to the document
4 rating determined for said rating scale for each document of the subset in
5 accordance to the document's size group classification.

1 11. The method of claim 10, wherein weights are applied to said determined
2 document ratings for said rating scale as follows:

Document size range in (bytes)	Weight
<500	1
500 - 999	4
1000 - 4999	7
5000 - 9999	10
>9999	13

3
1 12. The method of claim 9, wherein said determining of the collection rating
2 comprises further subdividing said subset of documents into a plurality of groups in
3 accordance with their determined ratings for said rating scale, and applying a weight
4 to the document rating determined for said rating scale for each document of the
5 subset in accordance to the document's rate group classification.

- 1 13. The method of claim 12, wherein weights are applied to said determined
2 document ratings for said rating scale as follows:

Determined document rating for said rating scale	Weight
0	-0.5
1	0.5
2	3
3	6

3

- 1 14. The method of claim 9, wherein said determining of the collection rating
2 comprises computing the collection rating for said rating scale as follows:

$$CR = \frac{\sum_{i,j} r_i w_j \log(N_{ij} + 1)}{\sum_{i,j} w_j \log(N_{ij} + 1)}$$

4 where CR is the collection rating for said rating scale;

5 r_i is the weight applied for document rating group i ;

6 w_j is the weight applied for document size group j ;

7 N_{ij} is the number of pages in the collection with document rating i and
8 having group sizes j for said rating scale.

- 1 15. The method of claim 9, wherein said first collection of documents are web
2 pages of a web site, and said first subset of documents are textual documents of
3 said web site.

- 1 16. A method of operation on one or more data processing machines, the
2 method comprising:

3 determining whether a first document collection comprises at least one
4 document linked to at least one other document of at least one other second
5 document collection;

6 determining a collection rating for a rating scale for each of said at least one
7 other second document collection if said first document collection is determined to
8 comprise at least one document linked to at least one other document of at least
9 one other second document collection;

10 determining whether said first document collection comprises at least one
11 document being linked by at least one other document of at least one other third
12 document collection;

13 determining a collection rating for said rating scale for each of said at least
14 one other third document collection if said first document collection is determined to
15 comprise at least one document linked by at least one other third document
16 collection; and

17 determining a link rating for said rating scale for said first document collection
18 based on either said determined collection rating or ratings for said rating scale for
19 said at least one other second document collection, or said determined collection
20 rating or ratings for said rating scale for said at least one other third document
21 collection, or both, depending on whether collection rating or ratings are determined
22 for said rating scale for said at least one other second document collection, said at
23 least one other third document collection or both.

1 17. The method of claim 16, wherein each of said determining of a collection
2 rating for said rating scale for each of said at least one other second or third
3 document collection comprises determining document ratings for said rating scale
4 for documents of the particular document collection, and sizes of the documents,

5 and determining the collection rating for the particular document collection based on
6 the determined document ratings and the determined sizes.

1 18. The method of claim 16, wherein said determining of a link rating comprises
2 summing said collection rating or ratings determined for said rating scale for said at
3 least one other second or third document collection, and determining the link rating
4 based on the result of said summing.

1 19. The method of claim 18, wherein said determining of the link rating based on
2 the result of said summing comprises determining the link rating based on the result
3 of said summing as follows:
4

The result of said summing (RS)	link rating
RS less than -2	- 1.0
RS greater than or equal to -2, but less than - 1	- 0.5
RS greater than or equal to -1, but less than or equal to - 0.5	0
RS greater than -0.5, but less than or equal to 1.5	0.5
RS greater than 1.5, but less than or equal to 3	1.0
RS greater than 3, but less than or equal to 4	1.5
RS greater than 4	2.0

5

1 20. An apparatus comprising:
2 storage medium having stored therein a plurality of programming instructions
3 designed to enable said apparatus to
4 determine a first collection rating for a first rating scale for contents of a
5 first document collection,
6 determine a first link rating for said first rating scale for contents linked to
7 or linked by contents of said first document collection, and
8 modify said first collection rating for said first rating scale for contents of
9 said first document collection based on said determined first link rating
10 for said first rating scale for contents linked to or linked by contents of
11 said first document collection; and
12 at least one processor coupled to the storage medium to execute the
13 programming instructions.

1 21. The apparatus of claim 20, wherein said programming instructions are
2 designed to enable the apparatus to perform said determining of a first collection
3 rating by determining said first collection rating based on document ratings of a first
4 subset of documents of said first collection of documents, and sizes of the
5 documents of the first subset of documents of the first document collection.

1 22. The apparatus of claim 21, wherein said first subset of documents of said first
2 document collection consists of first textual documents of said first document
3 collection.

1 23. The apparatus of claim 20, wherein said programming instructions are
2 designed to enable the apparatus to perform said determining of a first link rating by

3 determining at least a second collection rating for at least a second document
4 collection with documents linked to or linked by documents of said first document
5 collection, and determining said first link rating based on said determined at least a
6 second collection rating of said at least a second document collection.

1 24. The apparatus of claim 20, wherein said programming instructions are
2 designed to enable the apparatus to perform said modifying of the first collection
3 rating by replacing the determined first collection rating with said determined first link
4 rating.

1 25. The apparatus of claim 20, wherein said programming instructions are
2 designed to enable the apparatus to perform said modifying of the first collection
3 rating by adding said determined first link rating to the determined first collection
4 rating.

1 26. The apparatus of claim 20, wherein said programming instructions are
2 designed to enable the apparatus to perform said modifying of the first collection
3 rating by subtracting said determined first link rating from the determined first
4 collection rating.

1 27. The apparatus of claim 20, wherein said first document collection is a web
2 site, and said contents of said first document collection are web pages.

1 28. An apparatus comprising:
2 storage medium having stored therein a plurality of programming instructions
3 designed to enable said apparatus to

4 determine document ratings for a rating scale for a subset of documents
5 of a document collection,
6 determine sizes of the documents of said subset,
7 determine a collection rating for said rating scale for said document
8 collection based on said determined document ratings of said subset
9 of documents, and normalized by said determined sizes of said subset
10 of documents; and
11 at least one processor coupled to the storage medium to execute the
12 programming instructions.

1 29. The apparatus of claim 28, wherein said programming instructions are
2 designed to enable the apparatus to perform said determining of the collection rating
3 by further subdividing said subset of documents into a plurality of groups in
4 accordance with their determined sizes, and applying a weight to the document
5 rating determined for said rating scale for each document of the subset in
6 accordance to the document's size group classification.

1 30. The apparatus of claim 29, wherein said programming instructions are
2 designed to enable the apparatus to apply weights to said determined document
3 ratings for said rating scale as follows:

Document size range in (bytes)	Weight
<500	1
500 - 999	4
1000 - 4999	7
5000 - 9999	10
>9999	13

31. The apparatus of claim 28, wherein said programming instructions are designed to enable the apparatus to perform said determining of the collection rating by further subdividing said subset of documents into a plurality of groups in accordance with their determined ratings for said rating scale, and applying a weight to the document rating determined for said rating scale for each document of the subset in accordance to the document's rate group classification.

32. The apparatus of claim 31, wherein said programming instructions are designed to enable the apparatus to apply weights to said determined document ratings for said rating scale as follows:

Determined document rating for said rating scale	Weight
0	-0.5
1	0.5
2	3
3	6

33. The apparatus of claim 28, wherein said programming instructions are designed to enable the apparatus to perform said determining of the collection rating by computing the collection rating for said rating scale as follows:

$$CR = \frac{\sum_{i,j} r_i w_j \log(N_{ij} + 1)}{\sum_{i,j} w_j \log(N_{ij} + 1)}$$

where CR is the collection rating for said rating scale;

r_i is the weight applied for document rating group i ;

w_j is the weight applied for document size group j ;

8 N_{ij} is the number of pages in the collection with document rating i and
9 having group sizes j for said rating scale.

1 34. The apparatus of claim 28, wherein said first collection of documents are web
2 pages of a web site, and said first subset of documents are textual documents of
3 said web site.

1 35. An apparatus comprising:
2 storage medium having stored therein a plurality of programming instructions
3 designed to enable said apparatus to

4 determine whether a first document collection comprises at least one
5 document linked to at least one other document of at least one other
6 second document collection,

7 determine a collection rating for a rating scale for each of said at least one
8 other second document collection if said first document collection is
9 determined to comprise at least one document linked to at least one
10 other document of at least one other second document collection,

11 determine whether said first document collection comprises at least one
12 document being linked by at least one other document of at least one
13 other third document collection,

14 determine a collection rating for said rating scale for each of said at least
15 one other third document collection if said first document collection is
16 determined to comprise at least one document linked by at least one
17 other third document collection, and

18 determine a link rating for said rating scale for said first document
19 collection based on either said determined collection rating or ratings

20 for said rating scale for said at least one other second document
21 collection, or said determined collection rating or ratings for said rating
22 scale for said at least one other third document collection, or both,
23 depending on whether collection rating or ratings are determined for
24 said rating scale for said at least one other second document
25 collection, said at least one other third document collection or both;
26 and
27 at least one processor coupled to the storage medium to execute the
28 programming instructions.

36. The apparatus of claim 35, wherein said programming instructions are
designed to enable the apparatus to perform each of said determining of a collection
rating for said rating scale for each of said at least one other second or third
document collection by determining document ratings for said rating scale for
documents of the particular document collection, and sizes of the documents, and
determining the collection rating for the particular document collection based on the
determined document ratings and the determined sizes.

1 37. The apparatus of claim 35, wherein said programming instructions are
2 designed to enable the apparatus to perform said determining of a link rating by
3 summing said collection rating or ratings determined for said rating scale for said at
4 least one other second or third document collection, and determining the link rating
5 based on the result of said summing.

1 38. The apparatus of claim 37, wherein said programming instructions are
2 designed to enable the apparatus to perform said determining of the link rating

5

The result of said summing (RS)	link rating
RS less than -2	- 1.0
RS greater than or equal to -2, but less than - 1	- 0.5
RS greater than or equal to -1, but less than or equal to - 0.5	0
RS greater than -0.5, but less than or equal to 1.5	0.5
RS greater than 1.5, but less than or equal to 3	1.0
RS greater than 3, but less than or equal to 4	1.5
RS greater than 4	2.0